RSAA

RS ASSURANCE & ADVISORY

SOC 2 with AI in Scope

What changes in your examination, what doesn't, and the evidence your auditor will sample

WHITE PAPER · OCTOBER 2026

CONTENTS

What this paper covers

1	The criteria don't change; the evidence does	3
2	Bringing AI inside the system boundary	3
3	The system description is where gaps show first	4
4	The Common Criteria, criterion by criterion	6
5	When to add the other four categories	8
6	AI agents under SOC 2	9
7	Evidence, populations, and the Type II period	9
8	The exceptions auditors find most often	10
9	How RS Assurance & Advisory approaches the examination	11
A	Common Criteria: what changes with AI	12
B	Category selection and AI-specific focus	13
C	Agent controls mapped to SOC 2 criteria	13
D	System description checklist	14

EXECUTIVE SUMMARY

The short version

There is no "AI category" in SOC 2, and there is no separate AI criterion to meet. To pass a SOC 2 with AI in scope, the controls around your models, agents, and AI vendors must satisfy the same Trust Services Criteria as the rest of your system. The criteria do not change. What counts as sufficient evidence does.

That distinction is the whole paper. An organization that treats AI as a new compliance domain will build a parallel control set nobody tests. An organization that treats AI as a new population of systems, identities, changes, vendors, and logs inside an existing control environment will pass.

© 2026 RS Assurance & Advisory | rsassure.com

Four practical consequences follow. Your system description has to name the models, agents, and AI providers explicitly, and classify each provider as a vendor or a subservice organization. Your change management has to cover model versions, prompts, and system instructions, which most change processes currently do not. Every agent needs its own identity, because an agent running under a shared service account produces logs that cannot be attributed and an access review that cannot be performed. And your logs have to cover the full examination period, which is the single most common reason an otherwise sound AI control set fails a Type II.

This paper walks the Common Criteria with AI in scope, sets out when to add Processing Integrity, Confidentiality, Privacy, or Availability, and lists the exceptions we see most often. Mapping tables are included as appendices.

SECTION 1

The criteria don't change; the evidence does

SOC 2 examinations are performed against the Trust Services Criteria — the 2017 criteria, with points of focus revised in 2022. The Common Criteria, which make up the Security category, are required in every SOC 2. The other four categories — Availability, Processing Integrity, Confidentiality, and Privacy — are added when the service organization makes commitments that require them.

Nothing in that structure was written with large language models in mind, and nothing in it needed to be. CC6 asks whether logical access is restricted to authorized users; it does not care whether the authorized user is a person or an agent. CC8 asks whether changes are authorized, tested, and approved; it does not care whether the change is a code deployment or a model version upgrade. The criteria are deliberately technology-neutral, and that neutrality is why they absorbed AI without amendment.

What changes is the evidence. Prompt and action logs, model version history, agent identity provisioning records, and the AI provider's own SOC 2 report become populations your auditor samples. If those populations do not exist, or are incomplete, the control fails on evidence rather than on design.

> **VERIFY BEFORE YOU RELY ON THIS**
>
> AICPA guidance in this area continues to develop. The structural position above — that AI is tested against the existing criteria rather than against new ones — reflects the criteria as issued. Before relying on any specific statement about AICPA guidance in a proposal or a report, confirm the current position with your practitioner and against the latest AICPA publications.

SECTION 2

Bringing AI inside the system boundary

Scope in a SOC 2 is defined by the system that supports the service commitments you make to customers. The rule for AI follows directly: **any AI feature, hosted model, or agent that touches the in-scope system is in scope.** It does not

matter whether it was procured as AI, whether it is a headline feature, or whether it arrived as a toggle inside software you already owned.

Three categories cause most of the scoping arguments.

Embedded AI in existing tooling

A summarization feature inside your support platform, an assistant inside your code repository, a transcription service attached to meetings. These touch in-scope data and are frequently omitted because the underlying tool was already scoped and nobody re-examined it after the feature shipped.

Hosted model providers

When customer data leaves your environment for a model provider's API, that provider is performing part of your processing. The question is not whether to include them but how: as a vendor subject to your vendor management controls, or as a **subservice organization** whose controls are necessary for you to meet your commitments.

Agents

An agent is a privileged automated actor inside your system. It is in scope by the same logic that puts any service account in scope, and it attracts more scrutiny because it acts without a human in the loop.

> **VENDOR, OR SUBSERVICE ORGANIZATION?**
>
> The distinction drives real reporting consequences. If a model provider is a subservice organization, you choose the *carve-out* method — describing the complementary subservice organization controls you rely on, such as a commitment not to train on customer data — or the *inclusive* method, which brings their controls into your report and requires their cooperation. Carve-out is far more common for hosted-model providers. Either way, the decision belongs in the system description, and reading the provider's own SOC 2 report is part of the work, not an optional extra.

SECTION 3

The system description is where gaps show first

Before a single control is tested, the description of the system has to be fairly presented. In practice this is where AI gaps surface earliest, because a description written in the previous reporting period rarely mentions the capabilities that shipped since.

A description that holds up names every model, AI feature, and agent in the system. It identifies each AI provider and states whether it is treated as a vendor or a subservice organization. For carved-out providers, it sets out the complementary subservice organization controls you rely on. It lists the complementary user entity controls your customers must perform — reviewing AI output before acting on it is a common and important one. And it explains how AI output reaches customers, and where in that path a person reviews it.

© 2026 RS Assurance & Advisory | rsassure.com

Appendix D sets this out as a checklist. Work it before drafting controls, not after: a description that changes late forces control changes with it.

> **A DRAFTING NOTE THAT SAVES REWORK**
>
> Describe AI capability at the level of function, not product name. "A third-party large language model service used to draft customer-facing responses, with review by a support agent before sending" survives a vendor switch and a model upgrade. A description built around a specific model version will be inaccurate before the period ends, and an inaccurate description is a reportable problem in its own right.

© 2026 RS Assurance & Advisory | rsassure.com

SECTION 4

The Common Criteria, criterion by criterion

Every SOC 2 covers the Common Criteria. Below is where AI shows up in each, and what we would expect to test. The full mapping is in Appendix A.

CC1 — Control environment

Someone has to own AI governance by name. Management oversight of AI use needs to be visible in minutes or equivalent records, and accountability for AI-assisted work has to sit with a person. This is the criterion organizations most often satisfy with a policy and nothing else; a policy with no oversight record behind it is a design gap.

CC2 — Communication

AI use must be communicated internally to staff and externally to customers. Internally that means policy acknowledgments. Externally it means the system description and any customer-facing disclosure obligations in your contracts.

CC3 — Risk assessment

AI risks belong in the risk register explicitly: wrong or fabricated output, data leakage to a model provider, prompt injection, and AI-enabled fraud, which CC3.3 reaches directly through its fraud consideration. CC3.4 requires assessing changes that could significantly affect the system — adopting a new model, or materially changing how an existing one is used, is such a change.

CC4 — Monitoring

AI controls get evaluated like any others, and deficiencies get reported and tracked. There is no separate standard of care here, which is precisely the point.

CC5 — Control activities

An acceptable use policy, an approved-tool list, and defined review steps for AI output. The evidence that matters is not the policy document but proof that the review steps actually happen.

CC6 — Logical access

The densest criterion for AI. Access to models, prompts, training and retrieval data, and API keys must be restricted and reviewed. Each agent needs a distinct identity at least privilege. CC6.7 reaches data sent to model providers, since that is movement of information outside the boundary. CC6.8 reaches unapproved AI tools — the shadow AI problem — because an unsanctioned browser extension or note-taker is unauthorized software introduced into the environment.

© 2026 RS Assurance & Advisory | rsassure.com

CC7 — System operations

Monitoring for model drift, misuse, and unusual agent behavior, and AI scenarios represented in the incident response plan. If your IR plan has no playbook for "an agent took an action it should not have," that gap will be found.

CC8 — Change management

The criterion most organizations are least ready for. Model version changes, prompt and system-instruction changes, retraining, and agent permission changes are all changes to the system and all require authorization, testing, and approval. Prompts and system instructions are configuration that materially alters system behavior, and they are very often edited outside any change process at all.

CC9 — Risk mitigation

Hosted-model providers assessed as vendors or subservice organizations, with contract terms addressing training use of your data. A training opt-out that exists in a marketing page but not in your contract is not evidence.

© 2026 RS Assurance & Advisory | rsassure.com

SECTION 5

When to add the other four categories

Security is required. Add the others when your AI use makes a customer promise that needs testing — not to make the report look more comprehensive. Each added category brings criteria that will be tested and can produce exceptions.

Processing Integrity

Add when AI output affects customer results — calculations, decisions, or generated content delivered as part of the service. Testing focuses on output accuracy and completeness, input validation, human review, and error handling.

Privacy

Add when AI processes personal information. Testing focuses on notice, consent, use limitation, and automated decisions about people.

Confidentiality

Add when customer data goes into prompts, fine-tuning, or retrieval. Testing focuses on whether data trains models, retention terms, and deletion at contract end.

Availability

Add when customers depend on an AI feature staying up. Testing focuses on your dependency on the model provider and what happens when it fails.

Processing Integrity deserves a specific warning. It is the category customers most often ask for when AI generates something they rely on, and it is the hardest to satisfy, because it requires you to demonstrate that output is complete and accurate. If a generative feature produces content a customer acts on, think carefully about whether you can evidence that claim across a full period before committing to it.

© 2026 RS Assurance & Advisory | rsassure.com

SECTION 6

AI agents under SOC 2

An AI agent is any system that takes actions on its own rather than producing output for a person to review — sending email, moving money, changing records, modifying infrastructure. Five controls carry most of the weight, and each maps to criteria an auditor already tests.

1. **An agent inventory with a named owner.** Purpose, access, and accountable person for every agent. Tested under CC1.3 and CC6.1.
2. **Its own least-privilege identity.** Not a shared service account and never a person's credentials. Tested under CC6.1, CC6.2, and CC6.3 through provisioning and access review records.
3. **Human approval for high-impact actions.** Payments, deletions, external messages, production changes — with the approval evidenced, not merely configured. Tested under CC5.2, CC8.1, and PI1.3 where Processing Integrity is in scope.
4. **Complete, tamper-resistant logs.** Retention covering the full report period, with review evidence. Tested under CC4.1, CC7.2, and CC7.3.
5. **A kill switch you have tested.** A documented stop procedure with test records. Tested under CC7.4 and CC7.5.

The first two do most of the work, and they are sequential: without a distinct identity per agent, the inventory cannot be reconciled to anything, the access review has nothing to review, and the logs cannot attribute an action to an actor. An organization that fixes agent identity first finds that several other controls become testable almost for free.

SECTION 7

Evidence, populations, and the Type II period

A Type I examination reports on the suitability of control design at a point in time. A Type II reports on operating effectiveness across a period, and operating effectiveness is demonstrated by sampling from populations. Almost every AI-specific failure we see in a Type II is a population problem rather than a control problem.

The population has to be complete and accurate

When you hand an auditor a list of agent actions, model changes, or approvals, that list is information produced by the entity, and the auditor has to establish that it is complete and accurate before sampling it. An export that cannot be tied back to a system of record, or a query nobody can reproduce, undermines every test that depends on it. Decide early how each AI population will be generated and how its completeness will be demonstrated.

© 2026 RS Assurance & Advisory | rsassure.com

Logs have to span the whole period

This is the one that catches well-run organizations. Default log retention in many AI platforms and gateways is short. If your examination period is twelve months and your prompt or action logs retain thirty days, the population covers a fraction of the period and the control cannot be tested across it. Check retention settings against the period length before the period begins, not during fieldwork.

Approval needs an artifact

"Human approval required" configured in a tool is a design statement. The test is whether approvals happened and whether the approver had a basis to approve. A click-through with no record of what was reviewed will be treated as an exception, and reasonably so.

Prompt and model changes need tickets

If prompts and system instructions live in a configuration file or an admin console with no change record, there is no population to sample for CC8 as it applies to AI. Bringing prompts under version control and routing changes through the existing change process is usually the single cheapest remediation available.

SECTION 8

The exceptions auditors find most often

Shared credentials for agents

Agents running under shared service accounts or a person's credentials. Breaks attribution, access review, and least privilege at once.

Untracked model and prompt changes

Model upgrades or prompt changes pushed without a change ticket.

Unassessed AI providers

Providers never put through vendor assessment, or whose SOC 2 report was collected but never read.

Approval theater

"Approved" clicks with no evidence of real review.

Short log retention

Logs that don't cover the full Type II period.

Stale system description

AI capability in production that the description never mentions.

Every one of these is cheaper to fix before a period starts than during fieldwork, and four of the six are configuration changes rather than process changes. That is the argument for a readiness assessment ahead of a first Type II with AI in scope.

© 2026 RS Assurance & Advisory | rsassure.com

SECTION 9

How RS Assurance & Advisory approaches the examination

Scoping and system description

Identifying every AI feature, model, and agent that touches the system, classifying each provider, and drafting a description that will still be accurate at period end.

Type I and Type II examination

Senior-led fieldwork with practitioners who have tested these populations, not junior staff learning on your engagement.

Readiness assessment

Testing the AI control set against the criteria before the period begins, when log retention and agent identity are still configuration decisions.

Integrated reporting

Controls mapped across SOC 1, SOC 2, HIPAA, ISO 27001, ISO 42001, and CMMC, so evidence gathered once serves multiple examinations.

For organizations pursuing both a SOC 2 with AI in scope and ISO/IEC 42001 certification, the overlap is substantial — the agent controls, the AI inventory, the vendor assessments, and the change records serve both. Sequencing the two deliberately is usually worth more than running them independently.

© 2026 RS Assurance & Advisory | rsassure.com

APPENDIX A

Common Criteria: what changes with AI

Criteria	What changes with AI in scope	What auditors will test
CC1 Control environment	A named owner for AI governance; management oversight of AI use; accountability for AI-assisted work	AI policy, oversight minutes, role assignments
CC2 Communication	AI use communicated to staff and customers; the system description names models and agents	Policy acknowledgments; customer disclosures
CC3 Risk assessment	AI risks in the register: wrong output, data leakage, prompt injection, AI-enabled fraud (CC3.3), new models as significant changes (CC3.4)	Risk assessment that covers AI explicitly
CC4 Monitoring	AI controls evaluated like any other; deficiencies reported	Control evaluations; deficiency tracking
CC5 Control activities	Acceptable use policy, approved-tool list, and review steps for AI output	Policies and evidence they operate
CC6 Logical access	Access to models, prompts, data, and API keys; a distinct identity per agent; least privilege; data sent to model providers (CC6.7); unapproved AI tools (CC6.8)	Access lists and reviews; key management; egress and shadow AI controls
CC7 System operations	Monitoring for drift, misuse, and unusual agent behavior; AI incidents in the response plan	Alerts, log reviews, incident tickets
CC8 Change management	Model version changes, prompt and system-instruction changes, retraining, and agent permission changes all go through change control	Change tickets with approval and testing
CC9 Risk mitigation	Hosted-model providers assessed as vendors or subservice organizations; training opt-outs in contracts	Vendor assessments, contracts, their SOC 2 reports

© 2026 RS Assurance & Advisory | rsassure.com

APPENDIX B

Category selection and AI-specific focus

Category	Add it when	AI-specific focus
Processing Integrity (PI1)	AI output affects customer results: calculations, decisions, generated content	Output accuracy and completeness; input validation; human review of output; error handling
Confidentiality (C1)	Customer data goes into prompts, fine-tuning, or retrieval	Whether data trains models; retention; deletion at contract end
Privacy (P1–P8)	AI processes personal information	Notice, consent, use limits, automated decisions about people
Availability (A1)	Customers depend on an AI feature staying up	Dependency on the model provider; fallback when it fails

APPENDIX C

Agent controls mapped to SOC 2 criteria

Agent control	SOC 2 criteria	What to show
1. Agent inventory with an owner	CC1.3, CC6.1	Inventory with each agent's purpose, access, and named owner
2. Its own least-privilege identity	CC6.1, CC6.2, CC6.3	Distinct identity per agent; provisioning and access review records
3. Human approval for high-impact actions	CC5.2, CC5.3, CC8.1, PI1.3	Approval rules; samples of approved actions
4. Complete, tamper-resistant logs	CC4.1, CC7.2, CC7.3	Log configuration; retention covering the full report period; review evidence
5. A kill switch you've tested	CC7.4, CC7.5	Documented stop procedure; test records

© 2026 RS Assurance & Advisory | rsassure.com

APPENDIX D

System description checklist

- [ ] Every model, AI feature, and agent in the system, by name
- [ ] Each AI provider, identified as a vendor or a subservice organization
- [ ] For carved-out providers, the complementary subservice organization controls you rely on (for example, "does not train on customer data")
- [ ] Complementary user entity controls your customers must perform (for example, reviewing AI output before acting on it)
- [ ] How AI output reaches customers, and where a person reviews it

© 2026 RS Assurance & Advisory | rsassure.com

ABOUT

RS Assurance & Advisory

RS Assurance & Advisory is an assurance and advisory firm headquartered in Southlake, Texas. We work exclusively with seasoned practitioners — no junior staff on engagements — and our unified methodology maps controls across frameworks so your team answers each evidence request once.

Service lines: SOC 1, SOC 2, and SOC 3 · SOX · ISO 27001 · ISO 42001 · CMMC 2.0 · HITRUST · HIPAA · GDPR · Nacha ACH Rules Compliance · PCI DSS readiness

Richard Stevenson
Managing Partner, RS Assurance & Advisory
richards@rsassure.com · (903) 229-0341 · rsassure.com

This paper is an educational starting point. It is not a substitute for the AICPA Trust Services Criteria, for professional standards, or for your practitioner's judgment in a specific engagement. Criteria references are summarized and should be read against the criteria themselves. AICPA guidance on AI continues to develop; confirm the current position before relying on any statement here. *© 2026 RS Assurance & Advisory. All rights reserved.*